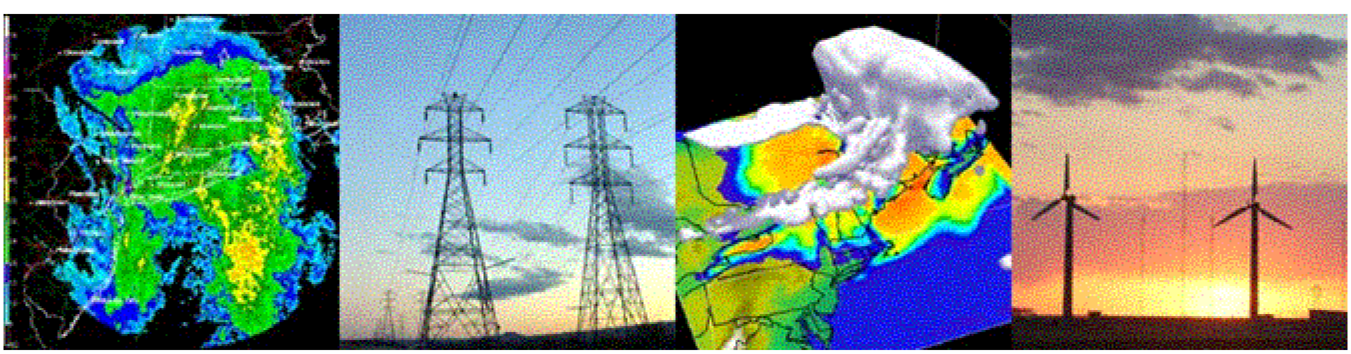
RUC Analysis-based CALMET Meteorological Data
for the State of North Dakota

Draft of Final Document

August 24, 2004



RUC Analysis-based CALMET Meteorological Data for the State of North Dakota

Dr. Dennis A. Moon

Abstract:
The process is described by which NOAA RUC analysis fields were used to derive meteorological input fields for the CALMET model. The ARPS Data Assimilation System (ADAS) assimilation tools developed by the University of Oklahoma were used to interpolate the RUC gridded data onto an MM5 grid, and assimilate surface observations. Adaptations of the ADAS code necessary to create the properly formatted files are discussed.

Introduction:
The CALMET/CALPUFF modeling system has been promulgated by the US EPA for regulatory permitting usage. CALMET is a diagnostic wind-field model used to generate three-dimensional meteorological fields as inputs to the CALPUFF dispersion modeling system. While CALMET is capable of generating these fields directly from a set of surface and upper air observations, such fields lack dynamic consistency as implied by the equations of motion. While CALMET does adjust the wind fields to conserve mass, conservation of momentum, thermodynamic energy, and water substance are not observed. Prognostic models take these quantities into account including non-linear and time derivative terms. As a consequence, the EPA encourages the use of meteorological fields from a prognostic model as first guess fields, by allowing the use of three years of meteorological data when using prognostic model fields as opposed to five years required when using traditional meteorological data. (US EPA 40 CFR Part 51, Appendix W) CALMET is then used to blend the observational data into background fields from the prognostic model.

The Mesoscale Model version 5 (MM5) modeling system from the National Center for Atmospheric Research (NCAR) and Pennsylvania State University is the most widely used mesoscale modeling system, having been deployed in a wide range of research and operational settings, see (http://www.mmm.ucar.edu/mm5/mm5-home.html). Meteorological fields from MM5 are frequently used as a source of first-guess fields for CALMET, and indeed the CALMET system includes software tools for the conversion of MM5 generated datasets into a CALMET-specific input text format. Use of the MM5 model allows for custom grid resolution and specification of physical parameterizations as required to represent the relevant flow features for a given location. In addition, the model includes a Four Dimensional Data Assimilation (FDDA) capability, which can be used to “nudge” the model towards the actual solution when applied to past cases for which observed data is available.

Another source of prognostic modeling data that has not been as widely considered within the air quality modeling community are the real-time modeling systems from the National Center for Environmental Predictions (NCEP). NCEP is charged with the task of development and operation of the suite of models used by the National Weather Service forecast operations. An integral step in the process of generating a model-based forecast is the preparation of the initial (or analysis) fields. This entails preparation of a three-dimensional, gridded analysis of the state of the atmosphere with as much accuracy as possible. This process involves using an assimilation system to blend the most recent observational data with first-guess fields from previous runs.

Of particular interest is the Rapid Update Cycle (RUC) system used to provide frequently updated short-term forecasts. The RUC cycle is unique among the NCEP forecast systems in that analyses are produced every hour, versus every six hours for the other models used for longer term forecasting, such as in the Eta and GFS modeling systems. The RUC cycle uses a process known as continuous assimilation, in which short, one hour forecast segments are interspersed with applications of the data assimilation process. This means that each hour, the one-hour forecast fields are corrected, based on the real-time data collected by The National Oceanographic and Atmospheric Administration (NOAA). These corrected fields serve as the starting point for the CALMET analysis described here. The one-hour analysis time step (not to be confused with the much smaller internal time step used within the RUC forecast model) is important in that the short duration for the forecast phase limits the magnitude of forecast errors that are an inevitable consequence of running any prognostic model. The model is never allowed to stray too far from the actual state of the atmosphere, within limits determined by observational sampling frequency, density, and accuracy. Frontal positions, for instance, can be reasonably well adjusted for in the assimilation process providing that the model frontal positions are close to those supported by the observations.

The one-hour analysis interval of the RUC system is also important for air quality modeling, in that CALMET requires its meteorological fields on a one-hour time step. Continuous assimilation cycles are complex and computationally expensive to run; however one can take advantage of the NOAA RUC cycle by using archives of the hourly analyses. In addition, NOAA, both NCEP and the Forecast System Laboratory (FSL), the developers of RUC, have access to a wealth of real-time observational data resources well beyond that available to private entities. As a result, their analyses include a more complete set of observations than could be assembled in an attempt to run a custom continuous assimilation cycle. Table 1 shows the data resources utilized in the RUC process.

Table 1: Data – Initialization of 40-km RUC at NCEP

Data Type	Number	Frequency	# obs in Study Area
Rawinsonde (inc. special obs)	80	/ 12h	2 (see figure 1)
NOAA 405 MHz profilers	31	/ 1h	0 (see figure 2)
VAD winds (WSR-88D radars)	110-130 / 1h		3 soundings (see figure 3 for more information)
Aircraft (ACARS) (V, temp)	1400-4500 / 1h		Some flight level winds/no vertical profiles
Surface/METAR - land (V, p_{sfc}, T, T_d)	1500-1700 / 1h		22 (see figure 4)
Buoy	100-150	/ 1h	NA
GOES precipitable water	1500-3000 / 1h		complete coverage
GOES cloud drift winds	1000-2500 / 1h		complete coverage
GOES cloud-top pressure	~10 km res / 1h		complete coverage
SSM/I precipitable water	1000-4000 / 6h		complete coverage
Ship reports	10s	/ 3h	NA
Reconnaissance dropwinsonde	a few	/ variable	NA

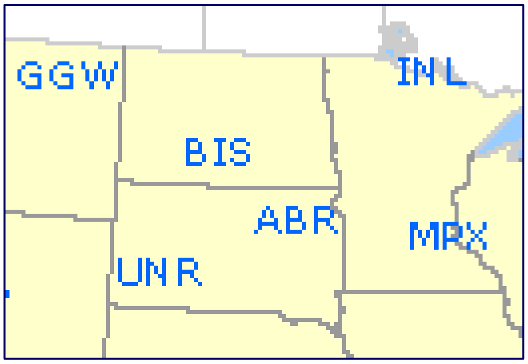


Figure 1. Location of radiosonde locations surrounding North Dakota

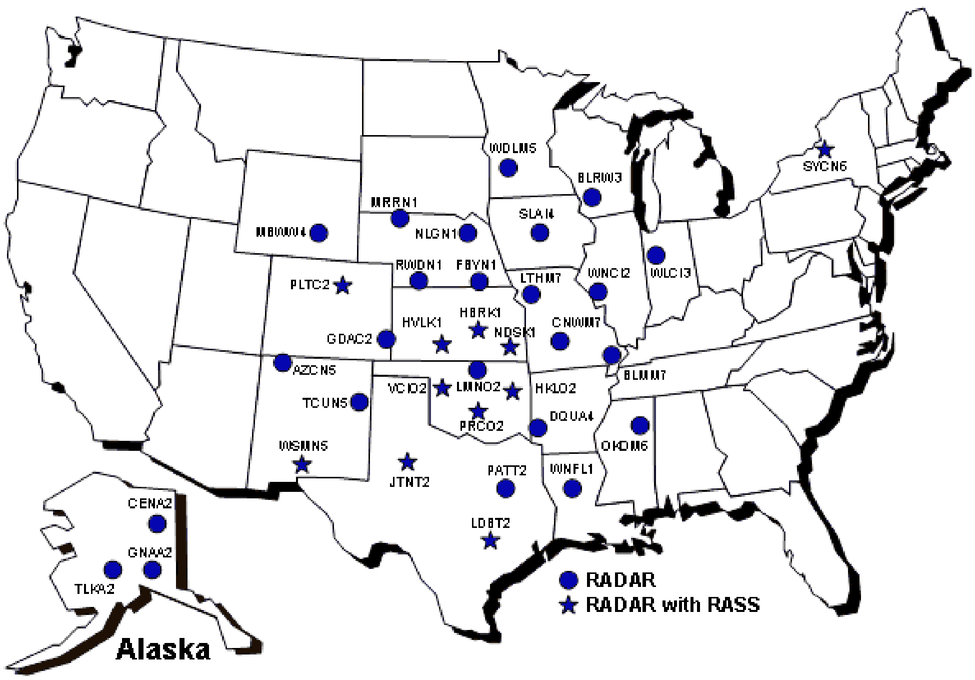


Figure 2. Locations of profiler and RASS stations

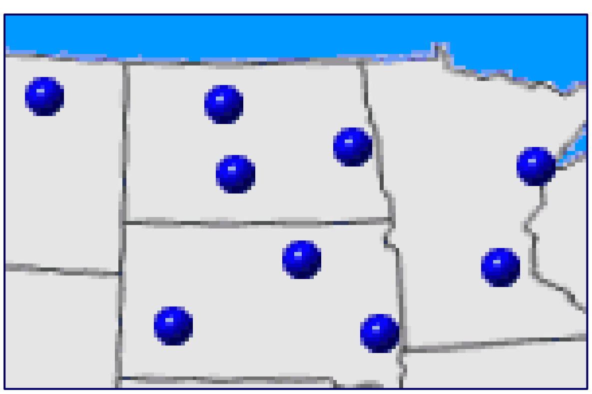


Figure 3. Locations of NWS-88D radars (normally 5-15 VAD wind levels available at each site, depending on atmospheric conditions).

The greatest value in the use of these analysis datasets as opposed to going back and re-simulating a time period using a mesoscale model lies in the application of the continuous assimilation process. Even when run in retrospective mode, model simulated fields will exhibit significant differences from observations. We often refer to this as "forecast" error and it is the result of imperfections in the finite representation of the continuous equations of motion, simplifying assumptions in the treatment of complex physical processes such as turbulence, radiation and cloud microphysics, errors in the representation of the soil and vegetative properties, and errors in the initial and boundary conditions supplied to the model which reflect the relatively low sampling density of the observational network. The net result is that while the flow patterns and resultant weather features are generally well simulated when considered from a pattern matching point of view, point time-series extracted from model simulations exhibit significant error, even when taken from what would be considered an excellent simulation. When simulating a long time-series the model must be periodically re-initialized from the archived data in order to eliminate model drift. Typically this might be done by restarting each day of the simulation, resulting effectively to the generation of a series of 24-hour forecasts.

Harrison (2003) discusses a comparison between 12-hour MM5 model forecasts and surface and upper air observations in the Pacific Northwest region over an 18-month period. The MM5 employed a grid resolution of 4 km. Model solutions were interpolated to the observation locations. Considering the surface (10m) wind results, he found the Root Mean Square (RMS) speed error to be 2.46 m/s, as compared to a mean wind speed of 2.55 m/s. The model surface direction showed a bias of –15.65 degrees, indicating a tendency for the model wind directions to be rotated clockwise or veered from the observations. The RMS direction error was a discouraging 73.65 degrees. He also calculated the coefficient of determination, R^2, for the matched model and observed wind vectors. The interpretation of this quantity is that a value of 1.0 indicates a perfect match, a value of 0.0 indicates a correlation between the two data sets no greater than that achieved by substituting the mean quantities as predictions (using climatology), and negative values indicate less correlation than the climatology. He reports a negative value of R^2 value at –1.25.

It should be noted that the Pacific Northwest weather patterns are strongly influenced by coastal and topographic effects, making it a very challenging area to model. However these values are not out of line with other sources. When comparing the NCEP Eta model forecasts against 50m winds observed at a midwestern wind turbine facility, WindLogics observed a <u>residual</u> mean absolute error (MAE) of 1.84 m/s after adjusting the model forecasts to remove the bias by applying a linear fit between the two sets of data. Brundage et al (2001) report on a comparison of 12-hour RUC model forecasts on a 40km grid against tower data. For a 10m AGL site near Golden Colorado, they report an RMS error for wind speed of 2.96 m/s. Comparing against 30m AGL tower measurements they report an RMS error of 2.49 m/s for a site in Southwest Minnesota.

The use of FDDA to nudge the simulation towards observed values can be of help in obtaining MM5 solutions that more closely match the observations, although in practice the use of FDDA can be problematic. Point nudging is a technique whereby the model experiences additional forcing terms at grid cells in the vicinity of an observational point which are proportional to the difference between the grid point value and the observed value. In this way the model solutions are "encouraged" to agree with the point values. This works well when in cases where the model solutions are inherently "close" to the observations, hence the apt choice of a name for the technique, nudging. In cases where the model solution diverges significantly from the observations the process of nudging can result in some very un-physical looking flow structures in the vicinity of the observations. This is particularly true in cases where the wind directions disagree. The problem is that the model winds at any point are the result of a complex set of interactions involving the mass and momentum fields over a large area. The process of nudging the model grid points near the observations cannot affect the large scale field from which the model winds result in any meaningful way.

Model first-guess fields derived from RUC archives have the advantage of being analyses rather than forecasts. That is, they reflect the level of error in the initial conditions used for the RUC model rather than the forecast errors. Therefore it would be expected that in many cases they should be suitable for use as high-quality first-guess fields for the CALMET modeling process. We present here a discussion of the processing steps used to convert the RUC data into a gridded form that is compatible with the CALMET system, and to re-introduce surface observation data through an assimilation process.

The CALMET modeling system data import processing programs for prognostic model fields are based on using the MM5 model as the data source. The data is made available to CALMET via a text file known as the MM5.DAT file, which contains a large number of header elements that describe the MM5 data set and grid parameters. Following the header is a sequence of data records covering multiple time steps. The process of getting the RUC data into CALMET comes down to making the transition from the RUC grid to the MM5 grid and writing the data into the MM5.DAT format. Note that the RUC uses a Lambert grid with different values for center longitude and "true" longitude. The true longitude is that at which the vertical grid lines are exactly aligned with true north and south. The center latitude and longitude mark the center of the grid. The MM5 map projection code makes the assumption that the center and true longitudes are the same value, thus it is not possible to create an MM5 grid that exactly matches the RUC grid.

WindLogics already had tools and procedures in place to handle this transition, in the form of the MM5 initialization system used to provide initial and boundary conditions for MM5 runs. The ADAS system from the Center for the Analysis and Prediction of Storms (CAPS) at the University of Oklahoma had previously been adapted to perform this task. ADAS also provides a sophisticated data assimilation capability used to blend observational data in with the RUC fields to produce a refined analysis. The only modifications required to the ADAS analysis system were those specifically related to the production of the MM5.DAT formatted file.

Procedures:

Data Characteristics:

The RUC analyses are in the form of GRIB formatted binary file, one file per hour. The data are on the NCEP 236 grid, with a horizontal resolution of 40 km. The data consists of a set of surface-type variables such as temperature and humidity at 2 m AGL, and winds at 10 m AGL, plus winds, temperature, and humidity on 19 pressure surfaces at an interval of 50 mb. The target grid in the ADAS system was a 10 km Lambert grid centered over western North Dakota. Figures 4 and 5 show the RUC and ADAS grids over the domain of interest. Also shown are surface metar stations within the study area. All of the stations shown are used by NCEP in the RUC process. The stations marked with dots show those with archived data available in the WindLogics archives for the process of re-introduction using ADAS.

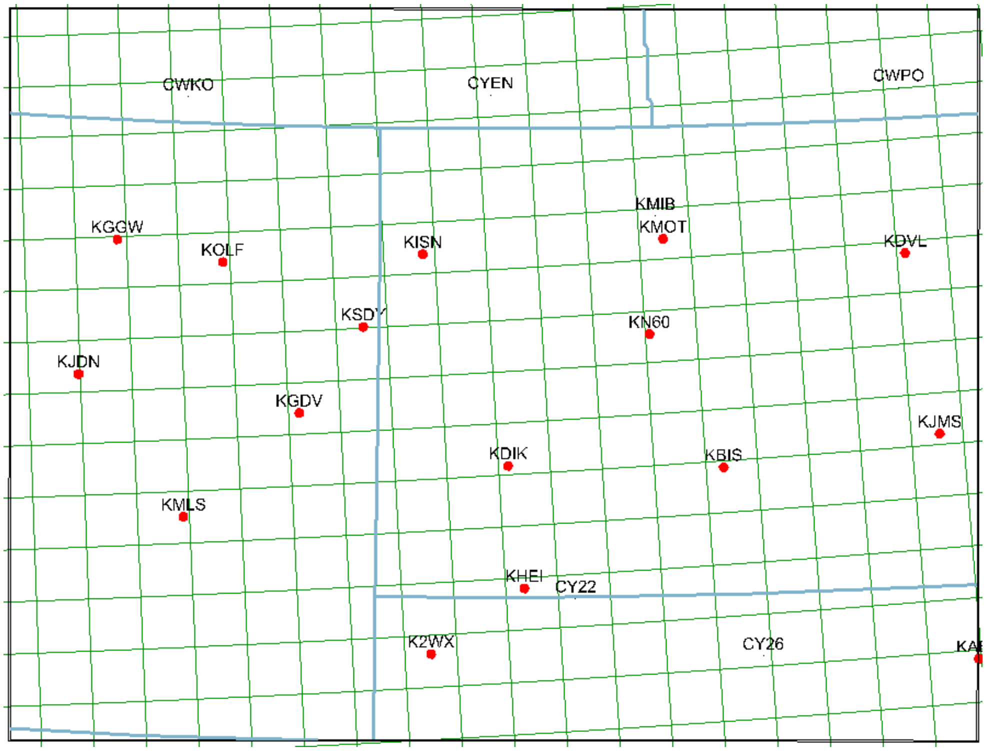


Figure 4. RUC 40 km grid over study area.

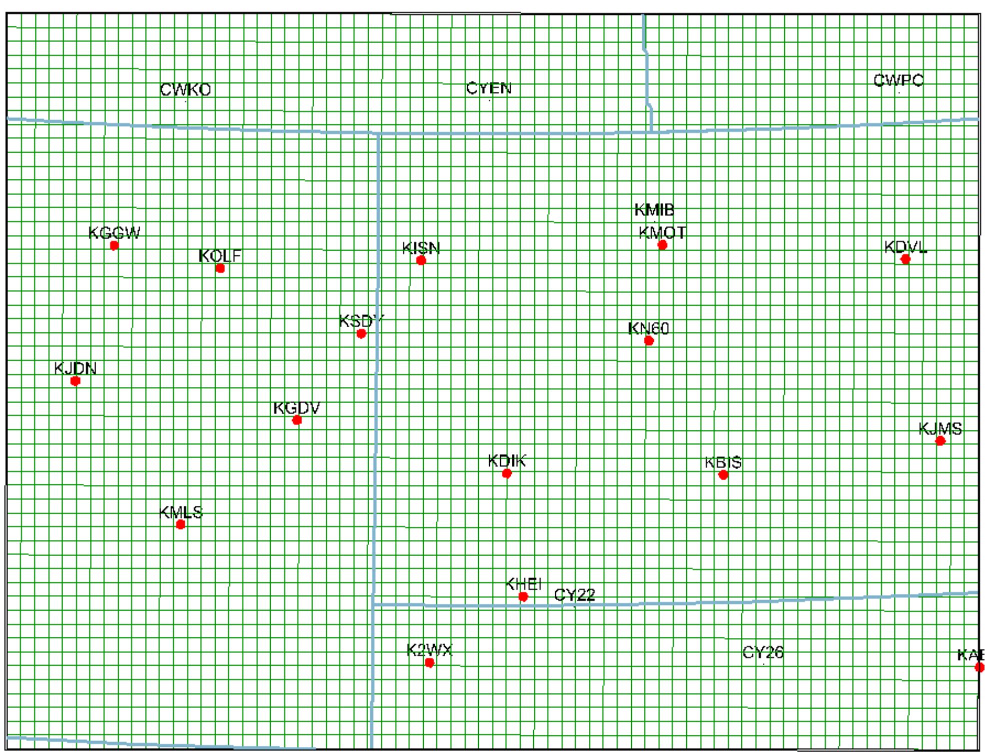


Figure 5. ADAS 10 km grid over study area

Processing:

The first step involves production of a terrain file for the target MM5 grid. This is done with the standard MM5 TERRAIN processor. WindLogics has modified the ADAS components to read standard MM5 terrain files. The next step involves reading in the GRIB-formatted RUC data and interpolating it to the target grid. This is done using the Ext2arps utility, a part of the ADAS system. This Fortran-based application is capable of accepting a wide variety of large-scale meteorological model data as input fields. The actions of Ext2arps are specified in a text file containing namelist formatted control variables, including the target grid specifications. In the North Dakota analysis the target grid has a cell spacing of 10km, compared to the 40km of the base RUC data, allowing for some additional resolution of finer-scale terrain influences on the flow field. However, since the Ext2arps process is based on interpolation, rather than flow dynamics as in a prognostic model, flows resulting from mesoscale processes operating on the finer terrain, such as thermally forced slope flows, will not be added into the flow fields if not present in the source RUC data. These kinds of effects are not considered to play a major role in the North Dakota grid. In addition, there are options in CALMET for inclusion of these effects via a parameterized diagnostic technique.

The native grid in the ADAS system is that of the ARPS mesoscale model, as shown in figure 6. The ARPS grid is similar to the MM5 grid, but differs in the grid stagger and vertical coordinate. Grid stagger refers to the convention used to define where different variables are located within a computational cell. The choice of grid stagger affects the level of accuracy in the discrete representation of the flow equations in the prognostic model. In the ARPS grid the scalar variables, temperature, pressure, and humidity are defined at cell centers, referred to as the “scalar” points. The X-component of the velocity is defined on the “eastern” faces of the cells, known as the “U” points. The Y-component of the velocity is defined on the “northern” faces of the cells, known as the “V” points. The vertical velocity component is defined on the top of the cells, the “W” points, at the same horizontal location as the scalar points. The vertical grid coordinate used by Ext2arps and ADAS is referred sigma-Z. The model levels have terrain following characteristics near the earth’s surface and gradually flatten out to constant height at higher altitudes. The cell spacing is stretched in the vertical, allowing for higher resolution near the earth’s surface to better represent the stronger vertical gradients present there. This is shown in Figure 7.

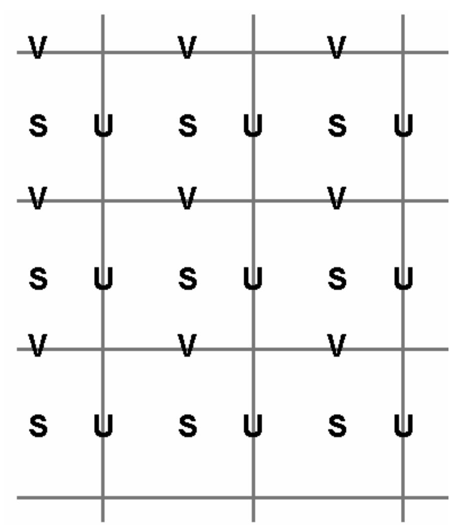


Figure 6. Grid stagger of the ARPS grid used by the ADAS system

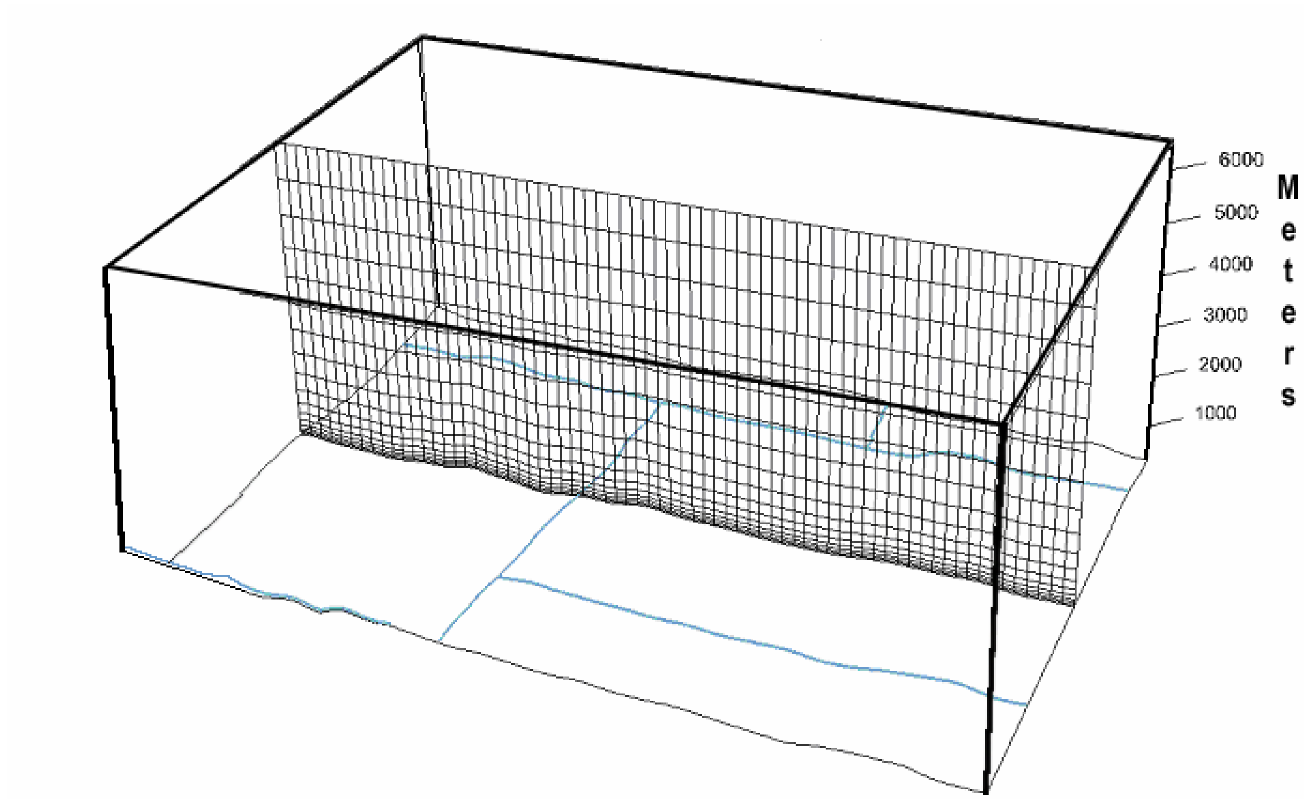


Figure 7. Vertical grid structure used by the ADAS system

Processing steps in Ext2arps include:

- Define the grid point locations for the target grid based on the user specifications.
- Find the Cartesian X, Y locations of the target grid points (scalar, U, and V points) in the external (RUC) grid map projection system. In that coordinate system the external grid rows and columns are orthogonal and lined up with the X and Y axes. This reduces map distortion errors and greatly simplifies the interpolation process.
- Interpolate the external grid quantities to the relevant target grid point locations (scalar, U, or V points), using bi-quadratic interpolation.
- Apply a single pass smoother to remove short-wave noise from the solution.
  - Simple 9-point smoother weighted as shown

    1 2 1
    2 4 2
    1 2 1
- Diagnose vertical velocity and adjust wind field to enforce conservation of mass.
- Apply an adjustment to the pressure variable to minimize vertical accelerations due to buoyancy when used to initialize a prognostic model.
- Write the analysis fields out as a binary file used to provide a first guess to the ADAS process.

Assimilation of observations and conversion of the data to the MM5 grid structure is done using the ADAS program. ADAS was developed for the purpose of creating initial conditions fields for mesoscale models. It uses the gridded fields developed in Ext2arps as its first guess fields and blends into them a potentially large variety of observation types including surface and upper air observations, aircraft data, Doppler radar reflectivity and radial velocity fields, infrared satellite radiance temperature fields, and visible satellite-derived albedo. In the current study, surface observations were the only observation type introduced through ADAS. Because of the large spacing between the radiosonde locations, relative to the RUC grid spacing, the RUC data was considered to be a fully resolved representation of the upper air data. The smaller spacing of the surface data introduces the possibility that the 40km RUC grid might not fully represent the flow pattern. The version of ADAS used by WindLogics has been modified from the original in two ways:

- The ability to interpolate from the native ARPS grid used by ADAS to the MM5 grid structure. In practice this involves accounting for the different grid stagger used by the two grid systems and interpolating in the vertical.
- The ability to write the MM5 gridded fields out to the MM5.DAT file used by CALMET.

Even if there are no observations to be assimilated, ADAS is still run to create the MM5.DAT file. Note that all assimilation is done on the observational increments rather than on the data values themselves. The observational increments are defined as the difference between the observed value and the first guess fields interpolated to the observation location. If the observation exactly matched the interpolated first-guess field the increment is zero. Positive values mean the observed value is greater than the background, and negative values indicate it is smaller. Interpolation to the grid is performed on the increments and the gridded increments are added to the first-guess fields to give the final results. This is standard practice in meteorological assimilation systems as it does an excellent job of retaining first-guess features in data poor regions.

The method used in ADAS to interpolate the discrete increments to the grid points is a variant on the technique often called “optimal interpolation” in meteorological circles. It is an efficient, iterative approach originally developed by the Norwegian meteorological service, known as the Bratseth method (Bratseth, 1986). The interpolation technique is “optimal” in the sense that based on statistical measures of the first-guess field and observational spatial variability; the resultant analysis error is minimized in a statistical sense. The optimal interpolation technique has been around since the 1960’s and is still in wide use at national meteorological centers. The improvements coming out of Bratseth’s work were primarily in achieving a computationally efficient iterative solution to the problem. This technique has several desirable attributes:

- When assimilating multiple observational data sets this method allows for expressing the differing degrees of confidence in the different data sources and uses this information to resolve conflicts between multiple imperfect data sources and arrive at a most probable solution.
- It handles common problems such as data clustering better than simple $1/R^n$ interpolation approaches.
- It employs a multi-pass approach, which allows for exclusion of individual datasets and modification of length parameters on a pass-by-pass basis.

Processing steps in ADAS include:

- Read in Ext2arps generated first-guess fields.
- Read in surface observations from a text file (optional).

- Quality control of the observations. The following tests are performed and data suspected of being bad is flagged and ignored in the assimilation process.
  - Climatological checks in which the data values are checked against climatological extremes
  - Temporal change checks to assure that the changes since the last hour are reasonable.
  - Spatial checks in which consistency with nearby observations is considered.
  - Comparison with the first guess field to ensure that the differences are not too large to be believable.
- For each observation point calculate the Cartesian X and Y location in the map projection system of the target grid.
- Interpolate the first guess gridded data to the observation points using tri-linear interpolation.
- Subtract the interpolated values from the observation values to form the discrete increments.
- For each variable, interpolate the discrete increments to the grid point locations to form the gridded increments.
- For each grid point add the gridded increment to the background fields.
- All analysis of wind components is done using grid relative U and V components to ensure they are orthogonal in the analysis coordinate system.
- Perform a vertical pressure gradient adjustment to minimize vertical accelerations due to buoyancy (applicable when used to initialize a prognostic model).
- Calculate the vertical velocity field and adjust to ensure mass conservation.
- Rotate the wind components from grid relative to true E-W and N-S representation.
- Interpolate the analysis quantities from the ARPS grid used by ADAS to the MM5 grid
  - Vertical interpolation is by way of linear interpolation from the ARPS sigma_z vertical levels to the MM5 sigma_p levels
  - Horizontal interpolation to the MM5 grid is simplified by the fact that both grids are based on the same specifications and share the same grid cell locations. The main issue to account for is differences in the way the variables are staggered.
    - Temperature, pressure, and humidity are already on the scalar points, which are conveniently co-located with the MM5 cross points, where the same variables are defined on the MM5 grid, so no interpolation is required.
    - The wind components need to be interpolated to the MM5 dot-points. For the X-velocity component, the MM5 dot point is midway between two ARPS U-points so linear interpolation is accomplished by simple averaging of the two surrounding U-points, as shown in figure 8. The same is true for the Y-component, which is easily interpolated from the two surrounding ARPS V-points.

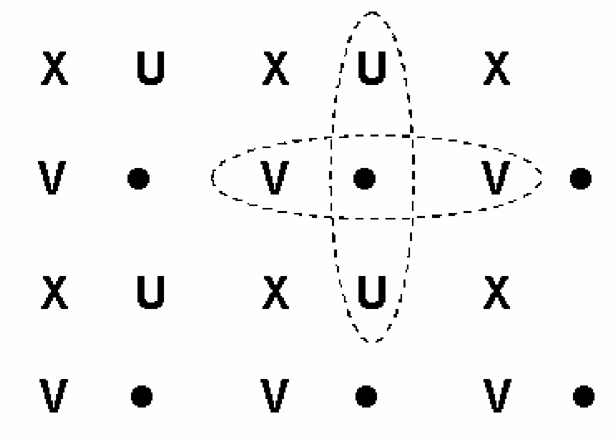


Figure 8. Interpolating velocity components from the ARPS grid to the MM5 grid dot points.

The next steps involve creation of the MM5.DAT file used by CALMET. The CALMET convention is that all quantities are read in from the MM5 dot-point locations, presumably to negate the need to interpolate the wind data.

- The wind X and Y components can be transferred into the MM5.DAT file as is, since they are already defined on the dot-points.
- Since temperature, pressure, humidity, and terrain height are all cross-point quantities they must be interpolated to the dot points before writing to the MM5.DAT file.
- Figure 9 illustrates how each dot-point is related to the four surrounding cross-points, making interpolation a matter of averaging the values from the four surrounding cross-points.

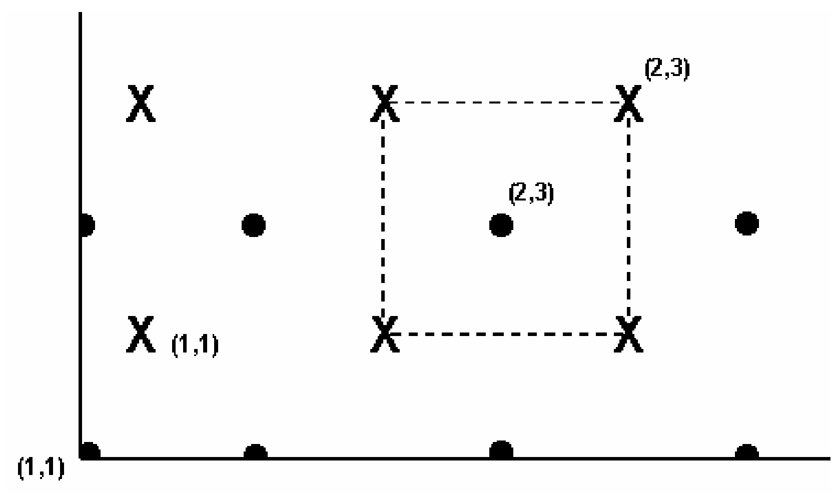


Figure 9. Interpolating to a dot-point from four surrounding cross-points**.**

Concluding Remarks:

The use of hourly RUC analyses from NOAA is for many situations an appealing option for providing first-guess fields to the CALMET modeling system. The task of extracting the RUC data and inserting it into the CALMET MM5.DAT format is made easier by the use of the ADAS mesoscale assimilation system, designed for use in initializing prognostic models. This flexible system handles the task of interpolating from the RUC grid to a user specified target grid using the Ext2arps utility, and enhancement of the data by assimilation of observations by way of the ADAS utility. Modifications were made to the system to allow the MM5 initialization data to be written in the MM5.DAT format.

References:

Bratseth, A. M., 1986: Statistical interpolation by means of successive corrections. *Tellus*, **38A**, 439-447.

Brewster, K. A., 1996: Application of a Bratseth analysis system including Doppler radar. Preprints, 15th Conference on Wea. Analysis and Forecasting, Norfolk VA, Amer. Meteor. Soc., Boston, 92-95.

Brundage, K.J., S.G. Benjamin, M.N. Schwartz, 2001: Wind energy forecasts and ensemble uncertainty from the RUC. Proceedings of AMS ninth conference on mesoscale processes.

Harrison, H, 2004: Is MM5 good enough for air quality models. http://www.atmos.washington.edu/~harrison/reports/mm5papr.pdf

About the Author

Dr. Dennis Moon, Chief Scientist, WindLogics Inc.

Dr. Moon has extensive experience in data analysis and numerical simulation of physical systems and is a recognized meteorologist and weather modeler. He directed the development of the Environmental WorkBench interactive 3D visualization application for environmental data, developing the set of user oriented functional requirements, and managing implementation in close collaboration with the software engineers. He has worked on a number of monitoring and simulation projects including the implementation of a real-time meteorological monitoring system for the Hanford Meteorological Station DOE site. He managed the development of a real-time wind field and toxic dispersion modeling system for the city of Cincinnati. He designed and provided scientific oversight for the deployment of sophisticated weather modeling system for the Israeli Air Force. He was the prime architect of the WindLogics wind characteristics analysis system that has been successfully used at a large number of sites across the US and the world. He has played a key role in a number of WindLogics project relating to simulation and analysis of wind patterns. He is a member of the American Meteorological Society.

For more Information, please contact:

WindLogics Inc., 1217 Bandana Blvd. N., St. Paul, MN 55108
Tel: 651.556.4200   Fax: 651.556.4210
E-mail: info@windlogics.com